



HAL
open science

Student assessment in an era of accountability

Richard Kitchen, Sarah Ridder, Joe Bolz

► **To cite this version:**

Richard Kitchen, Sarah Ridder, Joe Bolz. Student assessment in an era of accountability. CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education, Charles University in Prague, Faculty of Education; ERME, Feb 2015, Prague, Czech Republic. pp.1591-1596. hal-01287904

HAL Id: hal-01287904

<https://hal.science/hal-01287904>

Submitted on 14 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Student assessment in an era of accountability

Richard Kitchen, Sarah Ridder and Joe Bolz

University of Denver, Morgridge College of Education, Denver, USA, richard.kitchen@du.edu

In this research paper, the assessment practices of mathematics teachers at an urban high school in the USA who primarily taught racial and ethnic minority students and low-income students are described. We found that the teachers' assessment practices were largely influenced by the pressures to prepare students for success on the state's standardized test. For instance, teachers regularly used the language found on "the test" to classify students by their performance (i.e., Unsatisfactory), was regularly used by the teachers to label students, and was used in both assessment design and assigning student grades. Moreover, student performance on the test influenced how teachers viewed students, and consequently, how they viewed them in the assessment process. This is problematic given the long history in the USA of low-income, diverse students being denied access to challenging mathematics instruction.

Keywords: Assessment, diversity, urban education.

INTRODUCTION

The purpose of this research study is to examine how the assessment practices of mathematics teachers at a highly diverse urban high school in the United States of America (USA) were influenced by their students' performances on the state test ("the test"). At a time when large-scale assessments such as No Child Left Behind mandated tests (NCLB, 2001) [1] have been dominant in the USA, more research is needed to understand the impact of these tests on teachers' practices. It is important to understand how the test influences teachers' assessment practices because assessment is at the heart of what Ball and Forzani (2007) refer to as the "instructional dynamic."

The five teachers who participated in this study taught at Chavez High School [2] during the 2013–14 school year, a school situated in an urban school district in the state of Colorado. Chavez High served slightly more than 2,000 students in grades 9–12 and had a

highly diverse [3] student population. In 2013–14, the Free Reduced Lunch rate at the school was 75%; 64% of the Chavez High student body was Hispanic, 15% was African American, 13% was White, 4% was Asian American, and 1% was Native American or Native Hawaiian [4]. At the time this study was conducted, the standards-based Transitional Colorado Assessment Program (TCAP) assessment in mathematics was being administered to all grade 3–10 students in the state in the spring. The research question that we address in this paper is: How did the Colorado state mandated test in mathematics influence the assessment practices of mathematics teachers at Chavez High? Before we offer our research findings and address our research question, we provide reviews of the relevant research literature in classroom assessment and research pertaining to the mathematics education of low-income and diverse students. The research methodology used in this study is also described and we conclude with a brief discussion related to our research findings.

CLASSROOM ASSESSMENT

Classroom assessments are used to inform teachers, students and parents about student knowledge and understanding of mathematical concepts, processes and skills (Wiggins, 1993). Two ways of viewing assessment are assessment as an evaluative process focused on "an accounting of what is" (Webb, 1992, p. 663) and grade assignment (Kulm, 2013), and assessment as "the process of gathering evidence about a student's knowledge of, ability to use, and disposition toward mathematics..." (National Council of Teachers of Mathematics, 1989, p. 3). The former, summative assessment focuses on what students know at a given time (Guskey & Bailey, 2001). The latter, formative assessment, differs from summative assessment in that the focus is not just on summarizing students' learning, but on using information derived from assessments to inform instruction (Black & William, 2009). After more than a decade in which large-scale summative assessments such as No Child Left Behind

mandated tests (NCLB, 2001) have dominated the educational landscape in the U.S., we agree with Kulm (2013) that it is time to “reclaim the true meaning and purpose of enlightened mathematics assessment” (p. 4). Such assessment places a premium on using assessment to support teachers and the development of their instruction for the betterment of their students. For instance, Shepard (2000) called for assessments in which students actively make meaning of mathematical concepts by building on their previous knowledge and experiences and making connections to previous knowledge and new understandings.

With the introduction of the No Child Left Behind (NCLB) legislation and the testing that accompanies it (NCLB, 2001), educators and researchers have expressed a number of concerns about the impact that the introduction of high stakes, large-scale testing in the United States have had on teaching and learning (Nichols, Glass, & Berliner, 2006). For instance, students are reduced to test performers and “teachers find themselves using students to protect or help themselves... The marketplace mentality expands its reach” (Gergen & Dixon-Román, 2014, p. 8). “Teaching to the test” not only limits the content that is taught to what is tested, but also promotes superficial student learning that tends to be more skill focused and furthers student alienation toward school and learning (Nichols, Glass, & Berliner, 2006). Furthermore, instruction may be more targeted to those students who are close to meeting proficiency on the test and less instructional attention is paid to those who are far above or below meeting proficiency (Harlen, 2007). In a nutshell, high-stakes tests have had many negative consequences such as categorizing students as numbers to be compared to their peers, while limiting instruction to the detriment of students and their educational opportunities (Linn, 2000; Messick, 1995a; 1995b). Rather than holding schools and teachers accountable simply for student achievement on high-stakes tests, a range of student assessment information should be collected that will also inform teachers about how to improve their instruction (Harlen, 2007).

MATHEMATICS FOR LOW-INCOME AND DIVERSE STUDENTS

Globally, mathematics has served the historic role of sorting and stratifying students by race, ethnicity, and gender (Gerdes, 1988; Jurdak, 2014). In the U.S., white and Asian middle class and upper-middle class students have been privileged to have greater access to

challenging mathematics curriculum and instruction (DiME, 2007; Tate, 1995). Educational opportunities and access to such opportunities are influenced by where one lives, what Tate (2008) refers to as the “geography of opportunity” (p. 397). Massey (2009) contended that advantages and disadvantages procured from an individual’s socioeconomic status (SES) are both reinforced and compounded by geographic concentration. For instance, students from low-income communities attend schools in which pupil expenditures compare unfavorably to pupil expenditures in schools located in wealthy communities and achieve at lower levels than their wealthy counterparts (Payne & Biddle, 1999). Hoglebe and Tate (2012) found that algebra performance is also influenced by where students live; the SES of local communities is significantly related to students’ performance in algebra. Brynes and Miller (2007) found that SES has direct effects on mathematics achievement and indirect effects on both the opportunities students have to enroll in advanced mathematics classes in high school and on their propensity to take advantage of learning opportunities in mathematics.

In addition to poverty and SES, student access to a challenging standards-based mathematics education is influenced by race, ethnicity, and English language proficiency (DiME, 2007; Gutiérrez, 2008; Martin, 2013). A role that mathematics has played historically is to sort and stratify students by race, ethnicity, and gender (DiME, 2007; Gerdes, 1988). Specifically, white and Asian middle class and upper-middle class students have been privileged to have greater access to challenging mathematics curriculum and instruction (DiME, 2007; Tate, 1995). Schools that enroll large numbers of African American students often have disproportionately high numbers of remedial classes in mathematics in which instruction is focused on rote-learning and strategies that are intended to help students be successful on standardized tests (Davis & Martin, 2008; Lattimore, 2005). In response to NCLB (2001) and the demands to increase test scores, Davis and Martin (2008) argue that the preponderance of skills based instruction “[negatively] shape the lives of poor African American students in more significant ways than middle-class or affluent students” (p. 18). An extensive research base demonstrates that low academic expectations and lower pupil expenditures have historically been the norm for schools that serve students from low-income communities and racial and ethnic minority students (Ferguson, 1998; Knapp

& Woolverton, 1995). Given this research, it is not difficult to surmise that millions of low-income students of color are being denied access to instruction in which students regularly engage in mathematical reasoning and discourse to solve complex tasks (Davis & Martin, 2008; Kitchen, Burr, & Castellón, 2010; Téllez, Moschkovich, & Civil, 2011; Valero & Meaney, 2014).

RESEARCH METHODOLOGY

Beginning in the fall semester of 2013, we visited Chavez High 2–4 times a month for the duration of the 2013–14 school year. A school visit included two classroom observations of one of the five participating mathematics teachers on consecutive days in fall 2013 and again in spring 2014. A classroom observation consisted of videotaping the participating teacher teach a mathematics lesson as well as videotaping a group or groups of students who had provided consent to participate in the study. Every attempt was made to videotape in a manner that minimized interference in the mathematics lesson (e.g., the video camera was placed in a location in the classroom such as the back of the room so as not to block students' view of their teacher, the whiteboard, and any other instructional resources used by the teacher). An interview was conducted with each participating teacher immediately following the first or second observation. Interviews with individual teachers were 30–45 minutes in length. Two focus group interviews were also conducted in spring 2014. Four of the five participating teachers participated in the initial focus group interview and all five of the participating teachers attended the second focus group interview. An interview was conducted with the Chavez High principal in the spring of 2014 as well.

For the purposes of this study, the data analysed were all the interviews conducted that included the individual interviews conducted with participating teachers, the two focus group interviews and the interview with the Chavez High principal. The interview transcripts were analysed using interpretive methods (Erickson, 1986; Maxwell, 2005). Each interview was read as a whole, followed by a period of open coding to allow for the emergence of themes, and themes were then compared across interviews conducted. After a set of themes were obtained from the dataset, we searched for commonalities and differences across interviews conducted (Miles, Huberman, & Saldaña, 2013). We also sought both confirming and disconfirming evi-

dence by searching for supportive and non-supportive evidence (Erickson, 1986; Miles, Huberman, & Saldaña, 2013).

The five mathematics teachers at Chavez High who participated in this study were Ms. H, Ms. K, Ms. S, Mr. T, and Ms. V. All five teachers were chosen by the school's administration as among the best mathematics teachers at the school and were recommended for inclusion in this study. Ms. S is Hispanic and was the only teacher of color in the group of participating teachers.

HOW "THE TEST" INFLUENCED ASSESSMENT PRACTICES

In this study, we investigated how the Colorado state mandated test in mathematics influenced the assessment practices of the mathematics teachers at Chavez High. It became clear from the outset of this study that the teachers' assessment practices we report here were largely the result of administrative mandates to improve student achievement at the school. For example, teachers were frustrated that students were often allowed to retake assessments on which they had not performed well. Ms. H discussed a concern shared by her colleagues about how some students approached assessments; "I'm sick and tired of hearing that before I hand out the test, 'there is a re-test, right?'" (Ms. H, Faculty Focus Group Interview, May 19th, 2014). Teachers explained that they were under pressure from the school's administration to do everything they could to avoid failing students in their classes. Teachers knew that if 20% or more of their students failed a class, they would be called in for a meeting with a Chavez administrator (Faculty Focus Group Interview, May 19th, 2014). Thus, students were often allowed to take an assessment more than once to improve their grade. Many of the assessment practices that the teachers pursued, and discussed below such as "deployment" and Exit Tickets had been mandated by the Chavez administration.

Our first finding was that because of the intense focus at Chavez High on students' performance on the TCAP ("the test"), it was common for teachers and administrators to refer to students as Advanced Proficient or just Advanced, Proficient, Partially Proficient ("Bubble Students") or as Unsatisfactory (or just as "Unsats"). This language reflected not only how students had performed on the test, but had also become language that teachers had adopted to design assess-

ments and even assign students grades. The “Bubble Students” and “Unsats” were specifically earmarked for additional mathematics instruction as a means to change their status to “Proficient students.” We also found the “Bubble Student” or “Unsat” label persisted; it was well documented who these students were and teachers were well aware of who was a Bubble Student and who was an Unsat student. “Bubble students” were specifically targeted for supplemental instruction since these students were within reach of achieving Proficient status on the test.

Though students were classified as a “Bubble Student” or as an “Unsat” based on performance on the previous year’s test, teachers also used these categories to label student performance on summative and formative assessments that they used in their classrooms. For instance, mathematics teachers at Chavez High School engaged in what was referred to as “deployment.” Students who were evaluated as Partially Proficient or Unsatisfactory on a unit test were “deployed” to receive supplementary instruction on the mathematics unit just completed. After the 1–2 day deployment, students were administered a post-test. When discussing the impact of one such deployment, Ms. K reported that “77% of the Unsats moved up to Partially Proficient” (Ms. K, April 17, 2014) on one of the re-tests administered.

Teachers also discussed aspects of classroom assessments they designed using the language from the test. For instance, parts of unit tests used during deployments included “Unsat” items, or questions that generally required less of students (e.g., recall of mathematical vocabulary). The “Unsat” portion of an assessment was meant to provide “access points” for students (Ms. S, Focus Group Interview, April 17, 2014). By this, Ms. S meant that the content of the “Unsat” portion should be more elementary, focusing for example, on vocabulary items in geometry. Teachers explained that the majority of the test items on summative assessments and even formative assessments such as Exit Tickets were “Unsat” items (e.g., 70–80% of the items), while the remainder of the items on the assessment were more advanced. Mr. T explained that he generally included only “the naked” or skills based tasks on his Exit Tickets (Mr. T interview, 2-20-14). The content of mathematics assessments was clearly influenced by the content teachers perceived would be included on the test. Ms. S, for instance, discussed

modifying the curriculum she used because, in her opinion, it did not align well with the TCAP.

Finally, mathematics teachers at Chavez High used the test language categories to assign grades to their students. For instance, Ms. S used the classification language used on the test as part of her grading system, she did not assign grades based upon percentages: “I don’t have any numbers in my grade book. It’s a Partial, Unsat, or Advanced” (Ms. S interview, 10-24-13). For Ms. S, grades were assigned based upon how each student was performing, in a holistic manner, relative to language aligned with the test: “Unsat is they have some knowledge of some of the math that we did, so that’s about a D. Partial Proficient is C-ish. And Proficient is about a B because you’re doing what the standards are asking you to do. For students to earn an A, they have to take the math they’ve been doing and apply it to new problems that hasn’t been taught to them.” (Ms. S interview, 10-24-13).

In summary, we learned that the TCAP, Colorado’s standards-based test profoundly impacted teachers’ assessment practices, including their grading practices. Because of the strong focus by the Chavez High administration on student achievement on the test, and their goal of increasing student test scores, much of how teachers classified students and designed instruction was driven by the need to improve student achievement. Perhaps more importantly, student performance on the test greatly influenced how teachers viewed students, and consequently, how they viewed them in the assessment process. The language that was used on the test to classify students by their performance on it (i.e., Advanced Proficient or Unsatisfactory) was used as a means to label students and had become the taken-for-granted language teachers had adopted in many of their daily practices.

DISCUSSION

Few studies exist in the U.S. that examine how a state’s large-scale test impacts high school mathematics teachers’ assessment practices, specifically at a secondary school that serves primarily low-income and diverse students. An important contribution to the research literature is our finding that the test influenced the mathematics teachers at Chavez High to frequently use assessments in a summative manner and that they used their assessment results to categorize students vis-à-vis student performance cate-

gories used on the TCAP. Moreover, the mathematics teachers were largely concerned with their students, mostly low-income, students of color, deriving correct answers to skills based assessment tasks, rather than engaging them in solving rich mathematical tasks. This is problematic given the long history in the U.S. of low-income, diverse students being denied access to challenging mathematics instruction (Davis & Martin, 2008; Kitchen, DePree, Celedón-Pattichis, & Brinkerhoff, 2007). The intense focus on preparing students for the test, specifically students who have historically been denied access to a strong mathematics education program in the U.S., is continuing the destructive legacy of prioritizing low-level mathematics instruction for students from marginalized and oppressed communities (Davis & Martin, 2008; Lattimore, 2005). Finally, attaching a label such as “Unsat” to any student is unjust, particularly so when we consider that most Chavez High students were from historically marginalized and oppressed communities.

REFERENCES

- Ball, D.L., & Forzani, F.M. (2007). What makes education research “educational”? *Educational Researcher*, 36(9), 529–540.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31.
- Brynes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity-propensity analysis. *Contemporary Educational Psychology*, 32, 599–629.
- Davis, J., & Martin, D. B., (2008). Racism, assessment, and instructional practices: Implications for mathematics teachers of African American students. *Journal of Urban Mathematics Education*, 1(1), 10–34.
- Diversity in Mathematics Education (DiME) Center for Learning and Teaching. (2007). Culture, race, power and mathematics education. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 405–433). Charlotte, NC: Information Age.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York, NY: Macmillan.
- Ferguson, R. F. (1998). Teachers’ perceptions and expectations and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 273–317). Washington, DC: The Brookings Institution.
- Gerdes, P. (1988). On culture, geometrical thinking and mathematics education. *Educational Studies in Mathematics*, 19, 137–162.
- Gergen, K. J., & Dixon-Román, E. J. (2014). Social epistemology and the pragmatics of assessment. *Teachers College Record*, 116, 11. Retrieved at <http://www.tcrecord.org/libproxy.unm.edu/library/Issue.asp?volyear=2014&number=11&volume=116>.
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press.
- Gutiérrez, R. (2008). A “gap gazing” fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, 39, 357–364.
- Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation*, 33, 15–28.
- Hogrebe, M. C., & Tate, W. F. (2012). Place, poverty, and Algebra: A statewide comparative spatial analysis of variable relationships. *Journal of Mathematics Education at Teachers College*, 3, 12–24.
- Jurdak, M. (2014). Socio-economic and cultural mediators of mathematics achievement and between-school equity in mathematics education at the global level. *ZDM – The International Journal on Mathematics Education*, 46(7), 1025–1037.
- Kitchen, R. S., Burr, L., & Castellón, L. B. (2010). Cultivating a culturally affirming and empowering learning environment for Latino/a youth through formative assessment. In R. S. Kitchen & E. Silver (Eds.), *Assessing English language learners in mathematics* [A Research Monograph of TODOS: Mathematics for All], 2 (pp. 59–82). Washington, DC: National Education Association.
- Knapp, M. S., & Woolverton, S. (1995). Social class and schooling. In J. Banks & C. Banks (Eds.), *Handbook of research on multicultural education* (pp. 548–569). New York, NY: Macmillan.
- Lattimore, R. (2005). African American students’ perceptions of their preparation for a high-stakes mathematics test. *The Negro Educational Review*, 56(2&3), 135–146.
- Linn, R. L. (2000). Assessments and accountability. *Education Researcher*, 29(2), 4–15.
- Martin, D. B. (2013). Race, racial projects, and mathematics education. *Journal for Research in Mathematics Education*, 44, 316–333.
- Massey, D. S. (2009). The age of extremes: Concentrated affluence and poverty in the twenty-first century. In H. P. Hynes & R. Lopez (Eds.), *Urban health: Readings in the social, built, and physical environments of U.S. cities* (pp. 5–36). Sudbury, MA: Jones and Bartlett Publishers.

- Maxwell, J. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.
- Messick, S. L. (1995a). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5–8.
- Miles, M.B., Huberman, A.M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage.
- Nichols, S., Glass, G., & Berliner, D. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved from <http://epaa.asu.edu/ojs/article/view/72>.
- No Child Left Behind Act of 2001* (2001). Retrieved at <http://www.ed.gov/policy/elsec/leg/esea02/index.html>.
- Payne, K. J., & Biddle, B. J. (1999). Poor school funding, child poverty, and mathematics achievement. *Educational Researcher*, 28(6), 4–13.
- Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Tate, W. F. (1995). Economics, equity, and the national mathematics assessment: Are we creating a national tollroad? In W. G. Secada, E. Fennema, & L. B. Adajian (Eds.), *New directions for equity in mathematics education* (pp. 191–208). New York, NY: Cambridge University Press.
- Tate, W. F. (2008). "Geography of opportunity": Poverty, place, and educational outcomes. *Educational Researcher*, 37, 397–411.
- Téllez, K., Moschkovich, J. N., & Civil, M. (Eds.) (2011). *Latinos and mathematics education: Research on learning and teaching in classrooms and communities*. Charlotte, NC: Information Age Publishing.
- Valero, P., & Meaney, T. (2014). Trends in researching the socio-economic influences on mathematical achievement. *ZDM – The International Journal on Mathematics Education*, 46(7), 977–986.
- Wiggins, G.P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
3. "Diverse students" refers to students who are members of a racial or ethnic minority group and is synonymous with "students of color," a phrase commonly used in the U.S.
4. The ethnic/racial categories reported here are the categories used by the school district in which Chavez is located.

ENDNOTES

1. The No Child Left Behind Act of 2001 was passed by the U.S. Congress and requires states to develop assessments to be administered at the state-level. To receive federal funding, states must give these assessments to all students at select grade levels < http://en.wikipedia.org/wiki/No_Child_Left_Behind_Act>.

2. "Chavez High" is a pseudonym.